ADEPT2 – Resequencing, SNP Discovery, and Association Studies for Loblolly Pine

A.J. Eckert[1], J.M. Lee[1], J.L. Wegrzyn[1], K. Stevens[1], C. Langley[1], and D.B. Neale[2]

[1]Section of Evolution and Ecology, University of California at Davis, Davis, CA 95616,
[2]Department of Plant Sciences, University of California at Davis, Davis, CA 95616

The ADEPT2 project aims to infer associations between natural genetic variation and complex phenotypic traits for loblolly pine (*Pinus taeda* L.). This is a collaborative project between the University of California at Davis (UCD), North Carolina State University (NCSU), the University of Florida (UF), and Texas A&M University (TAMU). Each partner is taking the lead on a part of the overall project, with UCD conducting all discovery and genotyping of single nucleotide polymorphisms (SNPs). Association analyses will subsequently be conducted on combined data sets consisting of our SNP data and the phenotypic data gathered at our partner institutions.

The resequencing of 7,850 EST unigenes in a diversity panel of 18 loblolly pine individuals has recently been completed by Agencourt Biosciences. All sequence data were generated from haploid DNA samples harvested from megagametophyte tissue, deposited within the TreeGenes database, and pushed through our sequence analysis, alignment, and SNP discovery pipeline. The average number of SNPs identified per amplicon is estimated to be approximately five, thus yielding a rough estimate of the average scaled population mutation rate (Q_W) of 3.44. We have also tested all primer pairs for single samples collected from *P. radiata* D. Don, P. lambertiana Douglas, *Picea abies* (L.) Karst., *Pseudotsuga menziesii* (Mirbel) Franco, and *Sequoia sempervirens* (D. Don) Endl. Observed rates of primer success across these taxa follow phylogenetic expectations, with success rates of 84%, 30%, 13%, 10%, and 2%, respectively.

Identified SNPs will then be evaluated for overall quality and ability to convert to standard genotyping platforms (e.g., Illumina GoldenGate Assay) using our automated SNP calling pipline. A subset of those SNPs will be used to genotype three loblolly pine association populations for which phenotypic data are or will be available from our partner institutions. Phenotypic data are grouped into five categories: wood properties (UF), disease resistance (UF), drought tolerance (NCSU), gene expression (TAMU), and metabolomics (UCD). A rangewide analysis of population structure and historical demography of loblolly pine using data from all three genomes is also underway to test assumptions of association-based analyses (i.e., lack of population structure), to provide appropriate null models for neutrality testing for the resequencing data, and to describe the phylogeography of this economically important species.

Our data provide an initial foundation for analyzing the diversity and structure of the loblolly pine genome and for the dissection of its complex phenotypic traits into their respective genetic components. Work at UCD has also resulted in a suite of bioinformatic tools and data available to the larger forest genetics community. We believe that these tools will help to answer the increasingly difficult questions arising in loblolly pine genetics and genomics.